

ADMINISTRASI UJIAN NASIONAL (UN) DENGAN MENGGUNAKAN MODEL COMPUTERIZED ADAPTIVE TESTING (CAT)*)

THE NATIONAL EXAM ADMINISTRATION BY USING COMPUTERIZED ADAPTIVE TESTING (CAT) MODEL

Handaru Catu Bagus

Puspendik Balitbang Kemdikbud, Jl. Gunung Sahari 4 Jakarta Pusat

Email: handcab@gmail.com / handcab@yahoo.com

Abstract: *Assessment model that ignores individual variations ability may cause information to be un-optimally received. Model of computerized adaptive testing (CAT) can get over these weaknesses because the level of difficulty of the item is adjusted with the abilities of students. The purpose of this study is to analyze the effectiveness, efficiency and accuracy of CAT models when used as an alternative replacement of conventional assessment models in national examinations (UN). Methodology of this research was quantitative comparative. This research used population of student answers that follow the UN province of Yogyakarta in 2009 with the subjects of mathematics and physics. The results of this study showed that the number of items selected by the CAT model is less than PPT model; the ability is adjusted to the level of participants; and there is a significant correlation with the PPT model. Therefore, CAT model is more efficient in term of time because it has fewer items than PPT model. It is also effective because it is adjusted to the ability of participants yet has the same accuracy compared to the PPT model.*

Keywords: *computerized adaptive testing, national examination, computer, assessment, adaptive model, item response theory*

Abstrak: *Model penilaian yang mengabaikan kemampuan variasi individu menyebabkan informasi yang diterima tidak akan optimal. Model computerized adaptive testing (CAT) dapat mengatasi kelemahan ini karena tingkat kesukaran soal menyesuaikan dengan kemampuan penempuh didik. Tujuan dari penelitian ini adalah untuk menganalisis efektivitas, efisiensi, dan keakuratan model CAT apabila digunakan sebagai alternatif pengganti model penilaian konvensional dalam ujian nasional (UN). Metodologi penelitian adalah kuantitatif komparatif. Penelitian ini menggunakan data populasi dari jawaban penempuh didik yang mengikuti UN di propinsi Daerah Istimewa Yogyakarta (DIY) tahun 2009 dengan mata pelajaran matematika dan fisika. Hasil penelitian ini terlihat bahwa jumlah soal yang dipilih oleh model CAT lebih sedikit dibandingkan dengan model PPT dan soal tersebut menyesuaikan dengan tingkat kemampuan penempuh serta terdapat hubungan yang signifikan dengan model PPT. Oleh karena itu, model CAT lebih efisien dalam hal waktu karena jumlah soal lebih sedikit dibandingkan dengan model PPT, efektif karena menyesuaikan dengan kemampuan peserta dan memiliki keakuratan yang sama dibandingkan dengan model PPT.*

Kata kunci: *komputer, Computerized Adaptive Testing, Ujian Nasional, penilaian, model adaptif dan teori respon soal*

Pendahuluan

Pasal 1 ayat 1 PERMEN DIKNAS No. 75 tahun 2009 menyatakan bahwa Ujian Nasional yang selanjutnya disebut UN adalah kegiatan pengukuran dan penilaian kompetensi peserta didik secara nasional pada jenjang pendidikan dasar dan menengah. Model penilaian pendidikan yang selama ini diterapkan di

Indonesia adalah tes tertulis selanjutnya disebut *paper and pencil test* (PPT) atau model konvensional. Tes tersebut menggunakan desain tes yang sama untuk peserta didik dengan usia atau jenjang pendidikan yang sama. Praktek semacam ini didasarkan pada asumsi bahwa peserta didik dengan usia atau jenjang pendidikan yang sama mempunyai

*) Diterima tanggal 3 Pebruari 2012 - dikembalikan tanggal 20 Pebruari 2012 - disetujui tanggal 1 Maret 2012

kemampuan yang sama. Padahal dalam kenyataannya terdapat variasi kemampuan yang signifikan.

Dalam teknis Prosedur Operasi Standar (POS) UN tahun pelajaran 2009-2010 pada bab III tentang Bahan Ujian Nasional terlihat jelas bahwa paket tes yang diberikan kepada peserta didik mengabaikan variasi kemampuan individu, mengingat bahwa pada paket tes UN untuk satu rombongan belajar pada satu sekolah yang disiapkan, digandakan, dan diberikan kepada peserta didik dengan menggunakan desain paket tes yang setara.

Pengabaian variasi kemampuan individu dalam model penilaian kompetensi tersebut mempunyai kelemahan, yaitu informasi yang diberikan tidak optimal. Sebagai contoh, paket tes yang dikategorikan mudah dan diberikan kepada sekelompok peserta didik yang memiliki kemampuan tinggi akan memberikan informasi yang kurang berarti, karena hampir dapat dipastikan bahwa semua peserta didik akan menjawab benar pada paket tes tersebut. Sebaliknya, bila paket tes dengan kategori sulit diberikan kepada sekelompok peserta didik yang memiliki kemampuan rendah maka kemungkinan semua peserta didik memperoleh skor 0 atau minimal. Dengan demikian, dampak yang timbul adalah masalah keadilan dan informasi yang dihasilkan akan menjadi tidak akurat.

Model *adaptive testing* dapat mengatasi kelemahan model penilaian pendidikan yang berlangsung selama ini. Model ini memungkinkan penggunaan tes yang *on target*, yaitu tes yang tingkat kesulitan soalnya sesuai dengan kemampuan peserta didik. Penggunaan tes tersebut akan menghasilkan informasi yang optimal dan tes akan berhenti sesaat setelah informasi kemampuan peserta didik dapat diestimasi. Teknik estimasi pengukuran model *adaptive testing* menggunakan pendekatan *Item Response Theory* (IRT). Pendekatan IRT berorientasi pada item tes, dan tidak berorientasi pada tes (Hambleton, 1991). Oleh karena itu, dengan pendekatan IRT performa seseorang atau sekelompok orang dalam sebuah item dapat diramalkan.

Untuk mempermudah proses model *adaptive testing* dalam penentuan soal yang sesuai dengan informasi kemampuan peserta didik maka model ini dibantu oleh media komputer atau terkomputerisasi sehingga hasil yang diperoleh akan lebih cepat, efektif dan akurat dalam menghasilkan informasi yang

optimal. Berdasarkan hal tersebut di atas, dapat dipahami bahwa *Computerized Adaptive Testing* selanjutnya disebut CAT adalah *adaptive testing* karena proses penyajian tesnya dibantu oleh media komputer termasuk dalam pemilihan butir soal, penampilan butir soal hingga pengolahan hasil tes. Bunderson (1989) mencatat beberapa kelebihan dari CAT, antara lain: meningkatkan kontrol dalam menampilkan item, meningkatkan keamanan tes, memperkaya kemampuan tampilan, diperoleh skor yang sama dengan waktu yang lebih singkat, mengurangi *error of measurement*, meningkatkan penyekoran dan pelaporan.

Penelitian model CAT ini sangat penting dilakukan sebagai alternatif pengganti model penilaian, khususnya UN yang selama ini diterapkan di Indonesia. Selain itu, masalah kebocoran dan kecurangan yang selama ini terjadi dalam penyelenggaraan UN dapat diminimalisir.

Rumusan masalah dalam penelitian ini antara lain: 1) apakah model CAT lebih efisien dibandingkan dengan model PPT apabila diterapkan dalam UN?; 2) apakah model CAT lebih efektif dibandingkan dengan model PPT apabila diterapkan dalam UN?; 3) seberapa akuratkah model CAT apabila dibandingkan dengan model PPT?

Tujuan penelitian ini adalah: 1) menganalisis efisiensi model CAT dibandingkan dengan model PPT apabila model CAT diterapkan dalam UN; 2) menganalisis efektivitas model CAT dibandingkan dengan model PPT apabila model CAT diterapkan dalam UN; 3) menganalisis hubungan atau korelasi antara hasil yang menggunakan model CAT dibandingkan dengan model PPT. Hal ini untuk mengukur keakuratan model CAT.

Terdapat dua manfaat dalam penelitian ini adalah: 1) sebagai pengenalan konsep baru kepada dunia akademik berupa bahan acuan dalam model penilaian pendidikan; 2) sebagai bahan masukan bagi penentu kebijakan dalam menyempurnakan sistem penilaian pendidikan di Indonesia.

Kajian Teori

Penilaian pendidikan merupakan bagian dari proses belajar mengajar. Model CAT merupakan salah satu dari penilaian pendidikan. Oleh karena itu, dapat dipahami bahwa penilaian adalah serangkaian kegiatan untuk memperoleh, menganalisis, dan menafsirkan data tentang proses dan hasil belajar

peserta didik yang dilakukan secara sistematis dan berkesinambungan, sehingga menjadi informasi yang bermakna dalam pengambilan keputusan oleh pihak sekolah atau pengambil keputusan (<http://rbaryans.wordpress.com/2007/07/27/>).

Pemberian penilaian dalam bentuk tes atau ujian kepada penempuh tes dapat dilakukan dengan berbagai cara, mulai dengan cara konvensional, yaitu dengan menggunakan kertas (*paper-pencil test*), hingga pemanfaatan teknologi, seperti komputer. Berikut akan dijelaskan mengenai bentuk administrasi tes yang banyak dilakukan.

1. *Paper-Pencil Test*

Administrasi tes melalui *paper-pencil test* (PPT) melibatkan penggunaan lembaran kertas. Tes pertama kali biasanya dibuat dalam bentuk PPT. Bentuk administrasi ini merupakan administrasi konvensional karena semua penempuh tes menerima seperangkat item yang sama. Kelemahan PPT adalah kerahasiaan tes tidak terjaga karena dapat saja dibaca oleh orang yang tidak berwenang atau bertanggung jawab (Bunderson, 1989). Selain itu, karena harus memberikan semua item, diperlukan waktu pengadministrasian yang lebih lama. Penggunaan kertas menjadi masalah tersendiri, misalnya dibutuhkan ruang untuk menyimpan data tes. UN yang sekarang diselenggarakan menggunakan administrasi PPT.

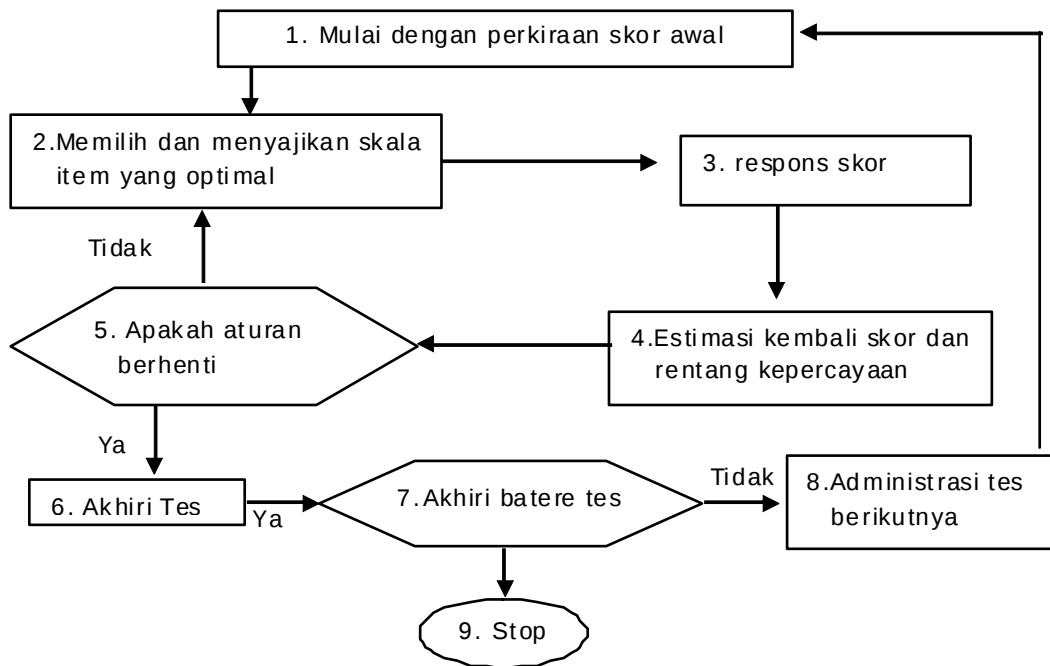
2. *Computerized Adaptive Testing*

Computerized Adaptive Testing (CAT) merupakan generasi kedua dari penggunaan komputer untuk pengesanan (Bunderson, 1989). Salah satu aplikasi dari pendekatan IRT adalah penggunaan CAT. Adaptif memiliki pengertian bahwa item yang diberikan sesuai dengan kemampuan setiap penempuh tes, sehingga setiap individu akan mendapatkan seperangkat item yang berbeda. Leung (2005) mengatakan bahwa dalam CAT seorang penempuh tes diberikan item yang dipilih berdasarkan kemampuannya yang diperkirakan (θ). Karena setiap individu mendapatkan seperangkat item yang sesuai dengan kemampuannya maka CAT termasuk dalam *tailored-testing*. Dengan demikian, CAT berbasis IRT biasanya berisi lebih sedikit item dibandingkan pengukuran PPT yang biasa atau konvensional (Embretson & Reise, 2000).

Dari penjelasan di atas, terlihat jelas bahwa CAT berbasis IRT biasanya berisi lebih sedikit item dibandingkan pengukuran PPT yang konvensional (Embretson & Reise, 2000). Hasil empirik juga dijelaskan oleh Olsen (dalam Bunderson, 1989) yang mencatat bahwa pada sebuah tes prestasi belajar hanya dibutuhkan 30% hingga 50% dari keseluruhan item tes untuk mencapai tingkat presisi yang sama dengan PPT. Dengan berkurangnya jumlah item yang diberikan kepada penempuh tes, maka secara langsung akan mengurangi jumlah waktu yang dibutuhkan untuk mengadministrasikan tes (Bunderson, 1989).

Konsep algoritma yang dipakai oleh CAT adalah sebagai berikut: apabila penempuh tes tidak dapat menjawab benar pada item yang diberikan maka komputer akan memberikan item yang memiliki derajat kesukaran yang lebih rendah. Sebaliknya, apabila penempuh tes dapat menjawab benar, item yang diberikan selanjutnya adalah item dengan kesukaran lebih tinggi. Gambar 1 memberikan bagan proses CAT yang dikemukakan oleh Wainer (1990).

Dalam mengaplikasikan sebuah tes ke dalam CAT yang berbasis IRT, ada beberapa hal yang harus diperhatikan. Embretson dan Reise (2000) memaparkan lima faktor harus diperhatikan dalam CAT. Pertama, **Item bank**. Tujuan dari CAT adalah untuk mengadministrasikan serangkaian item yang dapat memberikan informasi dan efisiensi yang maksimal untuk setiap penempuh tes. Untuk mewujudkan hal ini, penempuh tes yang berbeda akan menerima rangkaian item yang berbeda, dan skor mereka pada kemampuan laten diperkirakan berdasarkan respons mereka terhadap item-item yang berbeda tersebut. Dengan demikian, kapasitas untuk mewujudkan CAT bergantung pada sebuah item bank yang berisi sekumpulan besar item di mana parameter IRT dari setiap item telah diketahui. Tujuan pengesanan adalah untuk mengukur dengan baik keseluruhan rentang kemampuan, maka idealnya sebuah item bank berisi sejumlah item yang memiliki kemampuan daya beda tinggi dengan parameter kesukaran tersebar di antara rentang kemampuan. Ketika sebuah *item bank* memenuhi kriteria ini, seluruh penempuh tes dapat diadministrasikan pengujian dengan tepat dan mereka dapat diukur secara akurat. Kedua, **mengadministrasikan item pertama**. Apabila diasumsikan kemampuan penempuh tes dalam populasi terdistribusi secara



Gambar 1. Bagan Proses CAT

normal maka dapat dimulai dengan parameter kesukaran sebesar $-0,5$ hingga $0,5$. Apabila diperoleh informasi mengenai kemampuan penempuh tes dalam kontinum kemampuan maka informasi tersebut dapat digunakan untuk memilih tingkat kesulitan pada butir soal di awal. Rata-rata θ dari populasi penempuh tes dapat digunakan sebagai perkiraan kemampuan sehingga dapat menjadi optimal (Thissen & Mislevy, 1990). Ketiga, **pemberian skor**. Terdapat tiga metode utama untuk mengestimasi posisi penempuh tes dalam kontinum kemampuan, yaitu ML (*maximum likelihood*), MAP (*maximum a posteriori*), dan EAP (*expected a posteriori*). Beberapa peneliti tidak menganjurkan penggunaan informasi sebelumnya karena dapat berpotensi untuk mempengaruhi skor. Misalnya, apabila hanya sedikit item yang diadministrasikan maka tingkat kemampuan yang diestimasi akan tertarik ke arah nilai rata-rata dari distribusi awal. Untuk itu digunakan prosedur *step-size* untuk memberikan skor di tahapan awal CAT. Keempat, **pemilihan item berikutnya**. Pemilihan item berikutnya terkait dengan pemberian skor. Strategi yang dapat digunakan untuk memberikan item berikutnya adalah *maximum information* dan *minimum expected posterior standard deviation*, yang disebut juga *Bayesian estimation* (Thiessen & Mislevy, 1990). Pada *maximum information* dilakukan dengan memilih item pada setiap tahap yang memiliki nilai b mendekati perkiraan θ saat itu.

Kelima, **menghentikan tes**. Dalam CAT, setiap kali kemampuan penempuh tes diperkirakan kemampuannya berdasarkan respons terhadap item dan *standard error* diperkirakan kembali, komputer kemudian memilih item selanjutnya untuk diberikan. Ada dua kriteria untuk menghentikan administrasi CAT, yaitu *variable length* dan *fixed length*. Pada *variable length*, administrasi CAT berhenti ketika *standard error measurement* sudah mencapai batasan yang telah ditetapkan. Thissen dan Mislevy (1990) menyebut kriteria ini sebagai *target precision*. Penentuan *standard error*, menurut Hornke (2000), dengan *standard error* lebih kecil atau sama dengan $0,38$, akan sepadan dengan koefisien reliabilitas sebesar $0,85$. Di lain pihak, Blais dan Raiche (2002) menemukan apabila *standard error of measurement* lebih kecil atau sama dengan $0,40$ maka *standard error* dari tingkat kemampuan individu hanya berbeda sebesar $0,03$. Prosedur *fixed length* merupakan pemberhentian pengetesan apabila sejumlah item tertentu telah diadministrasikan. Thissen dan Mislevy (1990) menyebut kriteria ini sebagai *maximum number of items*. Kelebihannya adalah mudah untuk dilakukan dan penggunaan item dapat diperkirakan dengan tepat.

Penelitian ini menggunakan kerangka teori seperti disebutkan di atas sebagai dasar dalam metodologi proses penelitian model CAT. Oleh karena itu, lima faktor di atas menjadi sangat penting sebagai dasar model CAT dapat diaplikasikan.

Metodologi Penelitian

Metodologi penelitian ini menggunakan pendekatan kuantitatif dan bersifat komparatif. Data penelitian adalah jawaban atau respon siswa SMA yang mengikuti UN tahun 2009 di Provinsi Daerah Istimewa Yogyakarta (DIY). Hal demikian karena data respon UN yang dimiliki Provinsi DIY bervariasi sehingga mudah untuk diamati dan dianalisis. Sementara itu, penelitian difokuskan pada mata pelajaran matematika dan fisika, karena merupakan *core competence* peneliti. Sampel dipilih secara sistematis dari sejumlah populasi SMA yang mengikuti UN di Provinsi DIY. Cara pemilihannya adalah dengan mengurutkan data dari skor terendah hingga skor tertinggi, selanjutnya dipilih secara acak hingga total sampel menjadi empat ratus data yaitu terdiri dari seratus data dari persentil kurang dari 25, dua ratus data dipilih dari persentil antara 25 hingga 75 dan seratus data sisanya dari persentil lebih dari 75. Data sampel tersebut diolah dengan menggunakan administrasi model CAT dan dianalisis untuk melihat efisiensi, efektivitas dan akurasi dibandingkan dengan model penilaian konvensional atau model PPT.

Definisi Operasional

Sebelum menjelaskan metodologi penelitian lebih mendalam, terlebih dahulu akan dijelaskan definisi dari efektivitas, efisiensi, dan akurasi sebagai batasan penelitian ini. Menurut Hidayat (1986) efektivitas adalah suatu ukuran yang menyatakan seberapa jauh target (kuantitas, kualitas dan waktu) telah tercapai. Semakin besar presentase target yang dicapai, semakin tinggi efektivitasnya (<http://dansite.wordpress.com/2009/03/28/pengertian-efektifitas/>).

Sementara itu menurut SP.Hasibuan (1984) yang mengutip pernyataan H. Emerson, efisiensi adalah perbandingan yang terbaik antara *input* (masukan) dan *output* (<http://sitinurmala-mala.blogspot.com/2010/03/kerja-dan-kelelahan.html>). Dengan kata lain, efisiensi adalah sesuatu yang kita kerjakan berkaitan dengan hasil yang optimal dengan tidak membuang banyak waktu dalam proses pengerjaannya. Arti dari akurasi adalah seberapa dekat nilai hasil pengukuran dengan nilai sebenarnya (*true value*) atau nilai yang dianggap benar (*accepted value*).

Jika efisiensi, efektivitas, dan akurasi dikaitkan dalam konsep CAT seperti dijelaskan di kerangka

teoritis di atas, dapat dijelaskan bahwa efisiensi dipahami pada jumlah soal yang dikerjakan oleh penempuh tes yang menggunakan model CAT lebih sedikit dibandingkan dengan model PPT. Dengan demikian, model CAT akan lebih menghemat waktu dibandingkan dengan model PPT. Efektivitas pada pemilihan soal-soal yang diberikan kepada penempuh sesuai dengan kemampuan penempuhnya. Oleh karena itu, soal-soal yang tidak sesuai dengan penempuhnya tidak akan dipilih oleh CAT. Sementara itu, akurasi dapat dipahami walaupun, jumlah soal pada model CAT lebih sedikit namun informasi kemampuan penempuh tes tetap memiliki tingkat presisi yang sama dibandingkan dengan model PPT dengan asumsi bahwa skor dari hasil model PPT sebagai data sebenarnya. Keakuratan ini akan dibuktikan dengan mengkorelasikan antara skor hasil peserta didik dengan administrasi tes model CAT dan PPT.

Prosedur Administrasi Tes Model CAT

Metodologi penelitian ini akan dijelaskan pula tentang prosedur apabila administrasi tes menggunakan model CAT. Prosedur ini ada kaitannya dengan faktor yang dimiliki oleh CAT yang sudah dijelaskan pada kerangka teoritis di atas. Penelitian ini menggunakan prosedur sebagai berikut: 1) *item bank*, bank soal yang dipakai adalah informasi statistik soal pada UN tahun 2009 mata pelajaran matematika dan fisika dengan jumlah soal setiap pelajaran adalah 40 butir; 2) mengadministrasikan soal pertama, pemilihan soal pertama dipilih oleh komputer secara random, dan dipilih soal yang memiliki tingkat kesukaran sedang atau antara -0,5 hingga 0,5 dalam skala logit; 3) pemberian skor, model CAT dalam pemberian skor menggunakan konsep *maximum likelihood* dan prosedur *step-size* di tahapan awal CAT; 4) pemilihan item berikut, model CAT dalam pemilihan item berikut menggunakan *maximum information* dilakukan dengan memilih item pada setiap tahap yang memiliki nilai *b* mendekati perkiraan θ saat itu; 5) menghentikan tes, model CAT dalam menghentikan tes menggunakan *variable length*, administrasi CAT berhenti ketika *standard error measurement* sudah mencapai batasan kurang atau sama dengan 0,4.

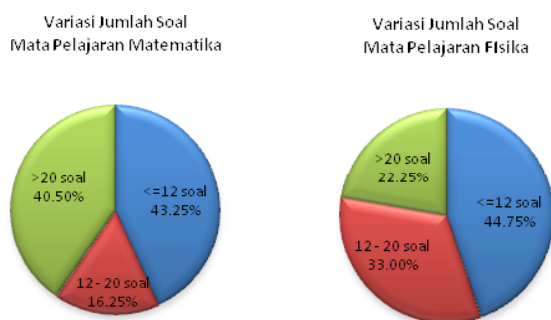
Analisis Hasil dan Bahasan

Dari hasil pengolahan penelitian, terdapat tiga hal yang dianalisis dan dibahas, semuanya memiliki kaitan

dengan tujuan dalam penelitian ini. Hal yang dibahas adalah 1) analisis efisiensi model CAT dibandingkan dengan model PPT; 2) analisis efektivitas model CAT dibandingkan dengan model PPT; 3) analisis hubungan atau korelasi antara hasil yang menggunakan model CAT dibandingkan dengan model PPT, hal ini untuk mengukur keakuratan model CAT.

Analisis dan Bahasan Efisiensi Model CAT Dibandingkan dengan PPT

Seperti telah dijelaskan di atas mengenai pengertian efisiensi jika dipakai dalam model CAT, yaitu pada jumlah soal yang dikerjakan oleh penempuh lebih sedikit dibandingkan dengan model PPT. Oleh karena itu, waktu yang dibutuhkan oleh penempuh akan lebih hemat. Apabila jumlah soal yang dikerjakan oleh penempuh dengan model PPT untuk setiap mata pelajaran matematika dan fisika adalah 40 soal dan waktu pengerjaannya 120 menit maka rata-rata satu soal adalah 3 menit. Dengan model CAT, penempuh tidak harus mengerjakan 40 soal. Jumlah soal yang dikerjakan bervariasi. Variasi tersebut dapat dilihat pada Gambar 2.



Sumber: Data Primer, diolah

Gambar 2. Variasi Jumlah Soal dengan Administrasi Tes Model CAT

Gambar 2 menunjukkan tiga variasi jumlah soal yaitu penempuh yang mengerjakan kurang dari 12 soal, antara 12 hingga 20 soal dan lebih dari 20 soal. Hasil persentase variasi jumlah soal di atas dihitung dari jumlah data sampelnya. Gambar di atas tampak bahwa administrasi tes dengan CAT lebih efisien dibandingkan dengan model PPT. Hal ini karena penempuh yang mengerjakan tes kurang dari 12 soal untuk mata pelajaran matematika adalah 43,25% sedangkan mata pelajaran fisika 44,75%. Dari keterangan di atas dapat dipahami bahwa penempuh yang mengerjakan tes 12 soal maka

model CAT akan lebih efisien 60 menit dibandingkan dengan model PPT.

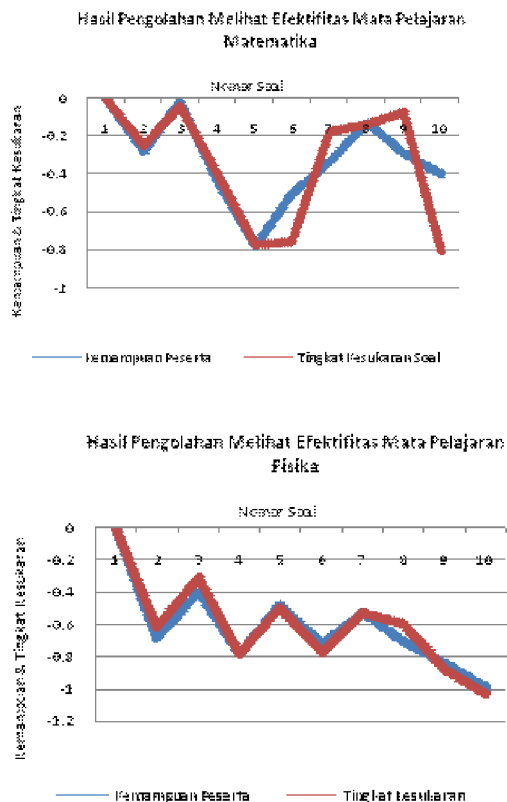
Sementara itu, penempuh yang mengerjakan tes lebih dari 20 soal untuk mata pelajaran matematika adalah 40,50% dan untuk fisika adalah 22,25%. Hal ini dapat dijelaskan bahwa banyaknya soal yang muncul ada kaitannya dengan sifat *item bank* dan konsistensi penempuh dalam menjawab soal. Apabila *item bank* yang dimiliki dalam administrasi tes model CAT memiliki sifat lebih menyebar dan dapat mewakili setiap skala kemampuan penempuh maka jumlah soal yang dikerjakan oleh penempuh tes akan sedikit dan tes akan cepat berhenti. Pengertian konsistensi penempuh dalam menjawab soal ada kaitannya dengan jawaban penempuh. Apabila penempuh menjawab satu soal maka model CAT akan mengestimasi kemampuan penempuh, dan kemampuan penempuh itu sebagai dasar dalam pemilihan soal berikutnya. Apabila soal berikutnya dipilih dengan tingkat kesukaran lebih rendah dari estimasi kemampuan penempuh maka penempuh harus menjawab benar dan sebaliknya apabila soal berikutnya dipilih dengan tingkat kesukaran yang lebih tinggi dari kemampuan penempuh maka penempuh harus menjawab salah. Penempuh terkadang tidak konsisten dalam menjawab soal, hal ini disebabkan penempuh yang tidak serius dalam mengikuti tes.

Selain efisiensi waktu pengerjaan tes dalam model CAT terdapat hal lain kaitannya dengan efisiensi yaitu tidak lagi menggunakan kertas. Hal ini karena model CAT menggunakan teknologi komputer, sehingga tidak ada lagi penggunaan paket tes. Oleh karena itu, biaya yang dibutuhkan untuk model CAT akan jauh lebih ekonomis jika dibandingkan dengan model PPT.

Analisis dan Bahasan Efektivitas Model CAT Dibandingkan dengan PPT

Efektivitas kaitannya dengan CAT seperti yang sudah dijelaskan pada metodologi di atas, yaitu pada pemilihan soal-soal yang diberikan kepada penempuh sesuai dengan kemampuan penempuhnya. Oleh karena itu, soal-soal yang tidak sesuai dengan penempuhnya tidak akan dipilih oleh model CAT. Hal ini jelas bahwa efektif ada kaitannya dengan efisien, yaitu jumlah soal yang dipilih akan disesuaikan dengan kemampuan penempuhnya. Jadi, setiap penempuh tes akan mendapat soal yang berbeda disesuaikan dengan kemampuan penempuhnya dan jumlah soal yang dikerjakan setiap penempuh tes tidak sama.

Gambar 3 adalah contoh hasil pengolahan salah satu penempuh dengan administrasi tes model CAT yang menggambarkan soal yang dipilih oleh model CAT disesuaikan dengan kemampuan penempuhnya.



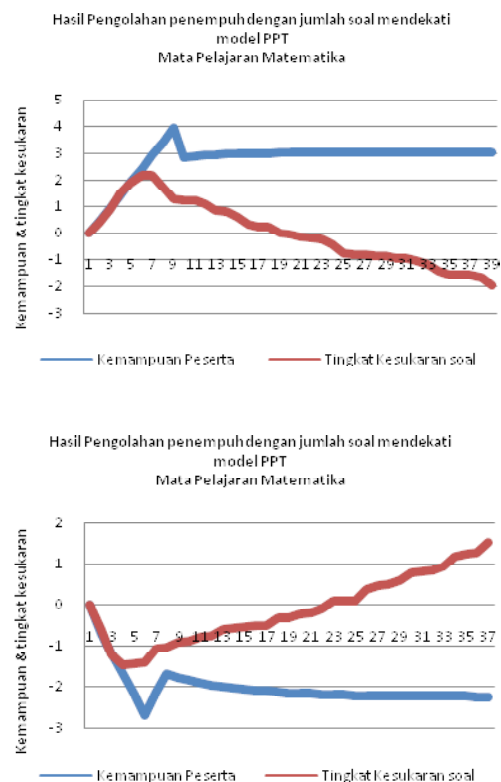
Sumber: Data Primer, diolah

Gambar 3. Hasil Efektivitas Model CAT

Gambar 3 tampak jelas bahwa soal yang dikerjakan oleh penempuh tes disesuaikan dengan kemampuannya, baik untuk mata pelajaran matematika maupun fisika. Khusus pada mata pelajaran matematika pada gambar 3, terdapat hasil yang jauh berbeda antara kemampuan penempuh dengan tingkat kesukaran soal pada nomor 6, 7, 9 dan 10. Hal ini karena kurang tersedianya *item bank* pada skala kemampuan penempuh. Soal nomor 6 contohnya, kemampuan penempuh ada pada level -0,51 dan seharusnya tingkat kesukaran soal yang dipilih oleh model CAT mendekati kemampuan penempuh tersebut, namun karena soal di *item bank* untuk tingkat kemampuan -0,51 tidak tersedia maka dipilihlah soal dengan tingkat kesukaran yang paling mendekati dengan kemampuan penempuh yaitu -0,76. Contoh yang paling ekstrim adalah nomor 9,

diketahui bahwa kemampuan penempuh ada pada level -0,4 dan seharusnya tingkat kesukaran soal yang dipilih mendekati kemampuan penempuh tersebut, namun karena soal di *item bank* untuk tingkat kemampuan -0,4 tidak tersedia maka model CAT memilih yang paling mendekati yaitu -0,81.

Berdasarkan analisis di atas jelas bahwa model CAT lebih efektif dibandingkan dengan model PPT, dan akan lebih efektif manakala *item bank* yang dimiliki oleh model CAT tersedia menyebar untuk setiap skala kemampuan penempuhnya. Jika *item bank* yang dimiliki tidak menyebar untuk setiap kemampuan penempuh maka penempuh tes dengan model CAT akan mengerjakan tes dengan jumlah soal mendekati atau sama dengan jumlah soal model PPT. Gambar 4 adalah contoh penempuh yang mengerjakan tes dengan jumlah soal mendekati model PPT namun dalam penelitian tidak banyak adalah 16% untuk pelajaran fisika dan 25% untuk mata pelajaran matematika.



Sumber : Data Primer, diolah

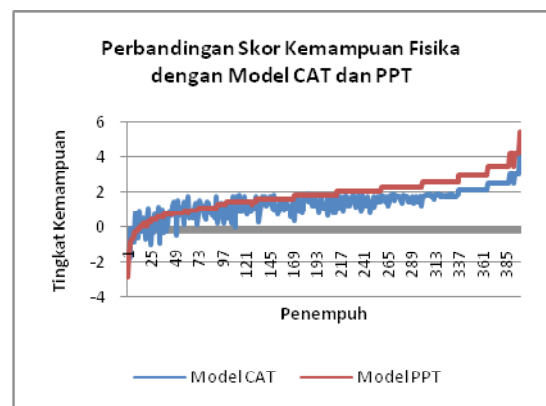
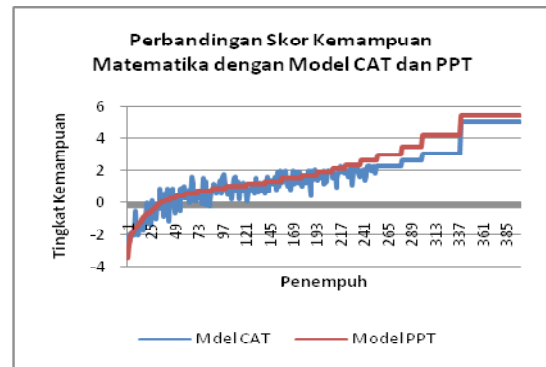
Gambar 4. Penempuh Tes dengan Jumlah Soal Mendekati Model PPT

Analisis dan Pembahasan Akurasi Model CAT Dibandingkan dengan PPT

Akurasi kaitannya dengan model CAT seperti yang sudah dijelaskan di atas dapat dipahami bahwa model CAT memiliki jumlah soal lebih sedikit namun informasi kemampuan penempuh tes tetap memiliki tingkat akurasi dan presisi yang sama dibandingkan dengan model PPT. Hal ini akan diuji dengan korelasi antara hasil skor model CAT dengan hasil skor model PPT. Konsep ini untuk melihat hubungan antara kedua model tersebut. Apabila signifikan maka terlihat jelas bahwa model CAT akurat.

Sebelum menjelaskan hasil korelasi antara model CAT dengan PPT, terlebih dahulu dijelaskan tentang data pengolahan skor kemampuan penempuh dengan menggunakan model CAT dan PPT. Hasil pengolahan terlihat bahwa skor minimal untuk mata pelajaran matematika dengan model CAT yaitu -2,44 logit dan -3,42 logit dengan model PPT. Skor maksimal dengan model CAT sebesar 5 logit dan 5,47 logit dengan model PPT. Untuk mata pelajaran fisika skor minimalnya sebesar -1,96 logit dengan model CAT dan -2,83 logit dengan model PPT, sedangkan skor maksimal dengan model CAT sebesar 5 logit dan 5,41 logit dengan model PPT. Perbandingan skor kemampuan dapat dilihat pada Gambar 5.

Gambar 5 jelas terlihat bahwa skor kemampuan penempuh dengan model CAT tidak ada perbedaan yang signifikan dibandingkan dengan model PPT. Pembuktian terdapat hubungan antara model CAT dan PPT perlu dilakukan analisis korelasi. Analisis korelasi penelitian ini menggunakan pendekatan *pearson correlation*, hal ini disebabkan skor kemampuan dengan model CAT dan PPT merupakan data interval dengan skala logit, dan skala ini dihitung dengan menggunakan pendekatan IRT. Hasil perhitungan korelasi *pearson* dapat dilihat bahwa terdapat hubungan yang signifikan antara skor kemampuan dengan model CAT dan PPT. Level signifikannya adalah 0,01 atau 1% dengan analisis *two tail*. Hal ini dapat dipahami bahwa dari 400 data sampel hanya terdapat 2 penempuh atau 0,5% memiliki skor kemampuan yang berbeda jauh antara model CAT dengan PPT.



Sumber : Data Primer, diolah

Gambar 5. Perbandingan Skor Kemampuan antara Model CAT dan PPT

Koefisien korelasi untuk pelajaran fisika adalah 0,894 dan untuk pelajaran matematika adalah 0,971. Hal ini dapat dipahami bahwa mata pelajaran fisika terdapat hubungan erat dan signifikan sebesar 0,894 antara model CAT dan PPT, begitu pula untuk mata pelajaran matematika terdapat hubungan erat dan signifikan sebesar 0,971 antara model CAT dan PPT. Nilai korelasi di atas mendekati nilai sempurna, hal ini dapat dipahami bahwa model CAT memiliki keakuratan dan presisi yang sama dengan model PPT.

Simpulan dan Saran

Simpulan

Dari penelitian ini menghasilkan tiga simpulan. Pertama, model CAT lebih efisien dibandingkan dengan model PPT. Hal ini dapat dipahami dari hasil kajian bahwa jumlah soal yang dikerjakan oleh penempuh dengan model CAT lebih sedikit dibandingkan dengan model PPT, sehingga waktu yang dibutuhkan lebih hemat. Model CAT dapat memiliki jumlah soal yang sama dengan model PPT

apabila *item bank* yang dimiliki tidak tersebar dalam variasi skala kemampuan penempeuh. Selain itu, konsistensi jawaban peserta akan mempengaruhi efisiensi model CAT. Kedua, model CAT lebih efektif dibandingkan dengan model PPT. Hal ini dapat dipahami dari hasil kajian bahwa soal yang muncul dan dikerjakan oleh penempeuh tes disesuaikan dengan kemampuan penempeuhnya, sehingga hasil yang akhir yang diperoleh lebih optimal dibandingkan dengan model PPT. Ketiga, hasil korelasi antara model CAT dan PPT menghasilkan koefisien yang positif dan signifikan. Hal ini dapat dipahami dari hasil kajian bahwa model CAT memiliki hubungan erat dengan model PPT, sehingga tingkat keakuratan model CAT sama dengan model PPT.

Saran

- 1) Pemerintah dalam hal ini Menteri Pendidikan dan Kebudayaan perlu memikirkan kebijakan UN dengan menggunakan model CAT sebagai alternatif pengganti model PPT;
- 2) *stake holder* pendidikan khususnya di sekolah, dapat mengaplikasikan model CAT sebagai model penilaian di sekolah;
- 3) bagi akademisi perlu dilakukan penelitian lebih lanjut dan mendalam, mengingat bahwa model CAT adalah model penilaian yang baru dan belum banyak diterapkan di Indonesia.

Pustaka Acuan

- Anonim. 2007. *Apa yang Harus Dilakukan Guru dalam Mengembangkan Silabus*. (<http://rbaryans.wordpress.com/2007/07/27/>, diunduh 28 Juli 2010).
- Anonim. 2009. (<http://dansite.wordpress.com/2009/03/pengertian-efektivitas/> diunduh 28 Juli 2010).
- Blais, J. & Raiche, G. 2002. Some Features of the sampling distribution of the ability estimate in computerized adaptive testing according to two stopping rules. Makalah disajikan pada 11th *International Objective Measurement Workshop*, New Orleans, April 2002.
- Bunderson, C.V., D. K. Inouye, and J.B. Olsen 1989. The Four Generations of Computerized Educational Measurement. Dalam Robert L. Linn. *Educational Measurement*. 3rd ed. New York: American Council on Education & Macmillan Publishing Company.
- Departemen Pendidikan Nasional. 2009. *Peraturan Menteri Pendidikan Nasional No. 75 Tahun 2009 Tentang Ujian Nasional SMP/MTs, SMPLB, SMA/MA, SMALB dan SMK Tahun Pelajaran 2009/2010*.
- Departemen Pendidikan Nasional. 2009. *Prosedur Operasi Standar (POS) UN SMP, MTs, SMPLB, SMA, MA, SMALB dan SMK Tahun Pelajaran 2009/2010*
- Embretson, S.E, and S.P.Reise. 2000. *Item Response Theory for Psychologist*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Hambleton, R.K., H. Swaminathan, and H.J.Rogers. 1991. *Fundamental of Item Response Theory*. Volume 2. California: Sage Publications, Inc.
- Hasibuan, S.P. 1984. *Manajemen Dasar dan Suatu Pengantar*. Jakarta: Haji Masagung.
- Hidayat. 1986. *Teori Efektivitas dalam Kinerja Karyawan*. Gajah Mada University Press. Yogyakarta
- Hornke, L.F. 2000. *Item Response Times in Computerized Adaptive Testing*. *Psicológica*. 21, 175-178.
- Kerja dan Keletihan. 2010 (<http://sitinurmala-mala.blogspot.com/2010/03/kerja-dan-keletihan.html>, diakses 28 Juli 2010)
- Leung, C., H.Chang, and K.Hau. 2005. Computerized Adaptive Testing: A Mixture Item Selection Approach for Constrained Situations. *British Journal of Mathematical & Statistical Psychology*, Nov 2005, 58, Proquest Psychology Journals, pp 239.
- Thissen, D., & R. J. Mislevy. 1990. Testing Algorithms. Dalam H. Wainer, N.J. Dorans, R. Flugher, & B.F. Green, *Computerized Adaptive Testing: a Primer*. New Jersey: Lawrence Erlbaum Associates, Publishers.
- Wainer, H. 1990. Introduction and History. Dalam H. Wainer, N.J. Dorans, R. Flugher, & B.F. Green. *Computerized Adaptive Testing: a Primer*. New Jersey: Lawrence Erlbaum Associates, Publishers.